WIND SPEED VERIFICATION OF ECMWF MONTHLY FORECASTS

BSC-ESS-2016-011

Nicola Cortesi, Nube González-Reviriego, Veronica Torralba and Francisco J. Doblas-Reyes

Earth Sciences Department

Barcelona Supercomputing Center - Centro Nacional de Supercomputación (BSC-CNS)

1 January 2016

TECHNICAL REPORT

Series: Earth Sciences (ES) Technical Report

A full list of ES Publications can be found on our website under:

https://earth.bsc.es/wiki/doku.php?id=library:external:technical_memoranda

Technical Report

BSC-ESS-2016-011

Summary

Validation of the past predictions of 10-metre wind speed of the monthly forecast system of the ECMWF revealed the existence of statistically significant skill over many areas of the world. This report aims to identify windows of opportunity, regions where monthly forecasts have a higher predictability than simple climatology for certain lead times, focusing on four key regions where wind power is generated: Europe, North Sea, Iberian Peninsula and North America. Results identified many windows of opportunity even beyond the first week of lead time, mainly during the winter half of the year.

Contents

Index of figures

1. Introduction

Monthly forecasting, also known as sub-seasonal forecasting, fills the gap between medium-range weather forecasting (up to two weeks) and seasonal forecasting (1-6 months). At monthly time scale, the atmospheric system has lost most of its memory from the initial conditions, which are relevant mainly during the first week; furthermore, the monthly scale is too short for the influence of the ocean state to differ significantly from its initial state and thus to beat persistence forecasts (Vitart et al. 2015). As a consequence, predicting hourly variations after the first week, particularly for a discontinuous variable such as wind speed, has little or no skill. However, there is opportunity for skilful predictions over longer time-averaging windows, for example by considering weekly averages, in which the unpredictable short-term fluctuations are reduced and predictive skill arises from the slow changes in the boundary forcing (Rodwell & Doblas-Reyes 2006). Weekly averages of monthly forecasts of surface wind speed, temperature and geopotential height have already demonstrated to produce statistically significant skill (Lynch et al. 2014, Weigel et al. 2008; Hudson et al. 2011). Forecast skill has been found during winter and over some European areas (particularly the United Kingdom) for weekly averaged wind speed over days 14-20 (Lynch et al. 2014). Thus, end users could gain potential economic value using the forecasts instead of the climatology to base their decisions.

A decade ago few operational meteorological services were producing sub-seasonal forecasts but progressively they have been incremented in number, partially due to initiatives such as the Sub-seasonal to Seasonal Prediction project (S2S), aimed to improve forecast skill, quantify its uncertainty and understanding systematic errors and bias at the subseasonal to seasonal timescale. It's focused also on identifying windows of opportunity for increased forecast skill, establishing multi-model database of ensemble of forecasts and promoting their uptake by operational centres and exploitation by the applications community. In the last decades forecast systems have slowly enhanced their skill, due to the increasing spatial resolution, the improved physical parametrizations (especially convection), better initial conditions and extended reforecast set. In recent years, many operational forecasting systems dedicated to sub-seasonal predictions have been implemented and now the majority of the Global Producing Centres (GPC) has a forecasting system designed to target the sub-seasonal time range. One of the more advanced sub-seasonal prediction systems is the multi-member ensemble monthly forecast system of the European Centre for Medium-range Weather Forecasts (ECMWF). In addition, other sub-seasonal prediction systems have also been developed by different institutions around the world, such as the Japan Meteorological Agency (JMA), the China Meteorological Administration (CMA), the National Centre for Environmental Prediction (NCEP/NCAR) or Météo-France (MF).

Many management decisions fall into the sub-seasonal scale, thus the predictability at this timescale promises to be of great economic and societal value (Robertson et al. 2015). Examples of sub-seasonal prediction applications can be found in many sectors, such as agriculture, where the end-users can support operational decision making on the timing of cultivating, irrigating, spraying and harvesting; insurance companies or financial institutions, where the end-users can also improve the decision making by trading commodities that are

impacted by weather; supermarket chains as the sales volume of certain products is dependent upon temperature or health services and disaster mitigation. In the case of renewable energy, and most specifically the wind energy sector can benefit from the sub-seasonal forecasts helping to stabilize energy costs and supply by improving scheduling and trading, maintenance scheduling, reducing curtailments and imbalance penalties, improving decisions about reserve energy sources, maximizing grid integration, and planning capacity commitments (Foley et al. 2012).

Inside of the wind energy context, the aim of this work is to examine the skill of the ECMWF monthly probabilistic forecast system in simulating the observed 10-metre wind speed at global scale, looking for so-called 'windows of opportunity', spatial regions where monthly forecasts have a higher skill than simple climatology for certain lead times. The first quality assessment of the European monthly wind speed forecasts was performed by Lynch et al. (2014) during winter, the season with the highest predictability, while Weigel et al. (2008) validated the monthly mean temperature at annual time scale. To the best our knowledge, this is the first attempt of validating seasonal 10-m wind speed forecasts both at world spatial scale and outside winter months.

Section 2 of this report outlines the data sources and methodology used. Section 3 examines the forecast quality assessment of the predictions using verification measures and focusing on the identification of windows of opportunities. In section 4, the main findings are discussed and general conclusions are drawn.

2. Data and methodology

2.1. Data and pre-processing

The version of the ECMWF monthly prediction system (ECMWF-MPS) employed in this study was publicly released in 2014. The ECMWF-MPS (Molteni et al. 2011) provides two forecasts per week with 51 members (simulations) each and forecasts 32 days long. Associated to the two forecasts of each week, one retrospective forecast (hindcast) is provided with 4 perturbed ensemble members, which have different initial conditions and/or physical parameters, for the past 20 years (1994-2013) and a spatial resolution of roughly 0.70°. For a full description of the ECMWF system, see Vitart et al. (2015). The reference dataset chosen for the quality assessment of the predictions was ERA-Interim reanalysis (Dee et al. 2011), with native spatial resolution of 0.75°. Forecast data was regridded to this resolution with a bilinear interpolation. See Figure 1 for an overview of the steps followed in data pre-processing.

In this work, we have only selected the 10m wind speed forecasts started on Thursday of each week and the analyses have been done for all the weeks through the year, being 4 or 5 different start dates available for each month (one for each week). For each start date, four lead times were selected averaging the 6-hourly raw forecast data (0000, 0600, 1200 and 1800 UTC) during weekly periods corresponding to days 5-11, 12-18, 19-25, 26-32. For each individual start date and lead time, daily anomalies were obtained at each grid point for both ERA-interim reanalysis and ECMWF-MPS forecasts separately, having as reference the climatology of the respective dataset for the period 1994-2013.

2.2. Methodology

To be able to develop a forecast quality assessment of the predictions, where the simultaneous predicted and observed values are compared, a time series of data long enough to provide strong verification measures is needed. Since the hindcast period is 20 years long (1994-2013), assessing the verification scores for a particular week only on the basis of such a short hindcast would not lead to robust scores. Instead, we obtained the verification measures for individual months using a novel technique in the bibliography where all the hindcasts in a month (one per week) were concatenated. Each verification score was obtained from a time series with a size of 80-100 pairs of observational and predicted values, result of concatenating the 4 or 5 weekly start dates in a month. This new approach provides a number of pairs sufficiently high to measure robust values of the skill scores (Wilks 2011).

A set of four verification measures, such as correlation of the ensemble mean (EnsCorr), fair rank probability skill score (FairRPSS), fair continuous rank probability skill score (FairCRPSS) and reliability diagrams, was applied in this study to assess the past performance of the predictions. All these measures were applied to the forecast anomalies rather than the absolute values, as there is a clear seasonality in wind speed. The first metric considered was EnsCorr, which is a deterministic score evaluating the predictions in terms of the temporal correlation coefficient between the anomalies of the ensemble mean and the observations,

on a grid point basis. The second metric is the commonly used probabilistic score RPSS, which measures the skill of the forecasts for categorical events (Wilks 2011). In this study, an enhanced version of the standard RPSS, known as FairRPSS (Ferro 2014), has been employed due to its advantage of not penalizing the intrinsic unreliability induced by small ensemble sizes. This property is favourable in the present work, since the ensemble size of the ECMWF-MPS is only of a few members. The FairRPSS applied here has been computed based on categorical tercile events. Values of FairRPSS below zero indicates that the predictions are unskilful, those equal to zero don't provide extra information than the climatology, and anything above zero is an improvement upon climatology, up to a maximum of 1, which indicates a 'perfect' forecast. The third metric, FairCRPSS (Ferro 2014), is also a probabilistic score but evaluate the skill of the full probability distribution instead of only three categories. As for the FairRPSS, values of FairCRPSS below 0 are defined as unskilful, those equal to 0 don't provide extra information than the climatology and anything above 0 is an improvement upon climatology, up to a maximum of 1, the "perfect" forecast (Joliffe & Stephenson 2011). To make inferences about the true value of the FairRPSS and the FairCRPSS, their p-value was estimated with a nonparametric approach, by means of the bootstrap method (Mason 2008), measuring 1000 skill scores obtained by resampling data with replacement.

Finally, reliability diagrams (Hartmann et al. 2002) of the forecasts examine the forecast frequency of the weekly average wind speed occurring in the lower, medium or upper tercile categories. They are simply graphs of the observed frequency of an event plotted against the forecast probability of the same event (in this case, the probability to belong to one of the tercile categories). This effectively tells the user how often (as a percentage) a forecast probability actually occurred and allows identifying any conditional or unconditional bias that may be exhibited by the forecasts. A perfect forecast system will result in forecast probabilities equal to the observed ones that correspond to the diagonal line of the reliability diagram.

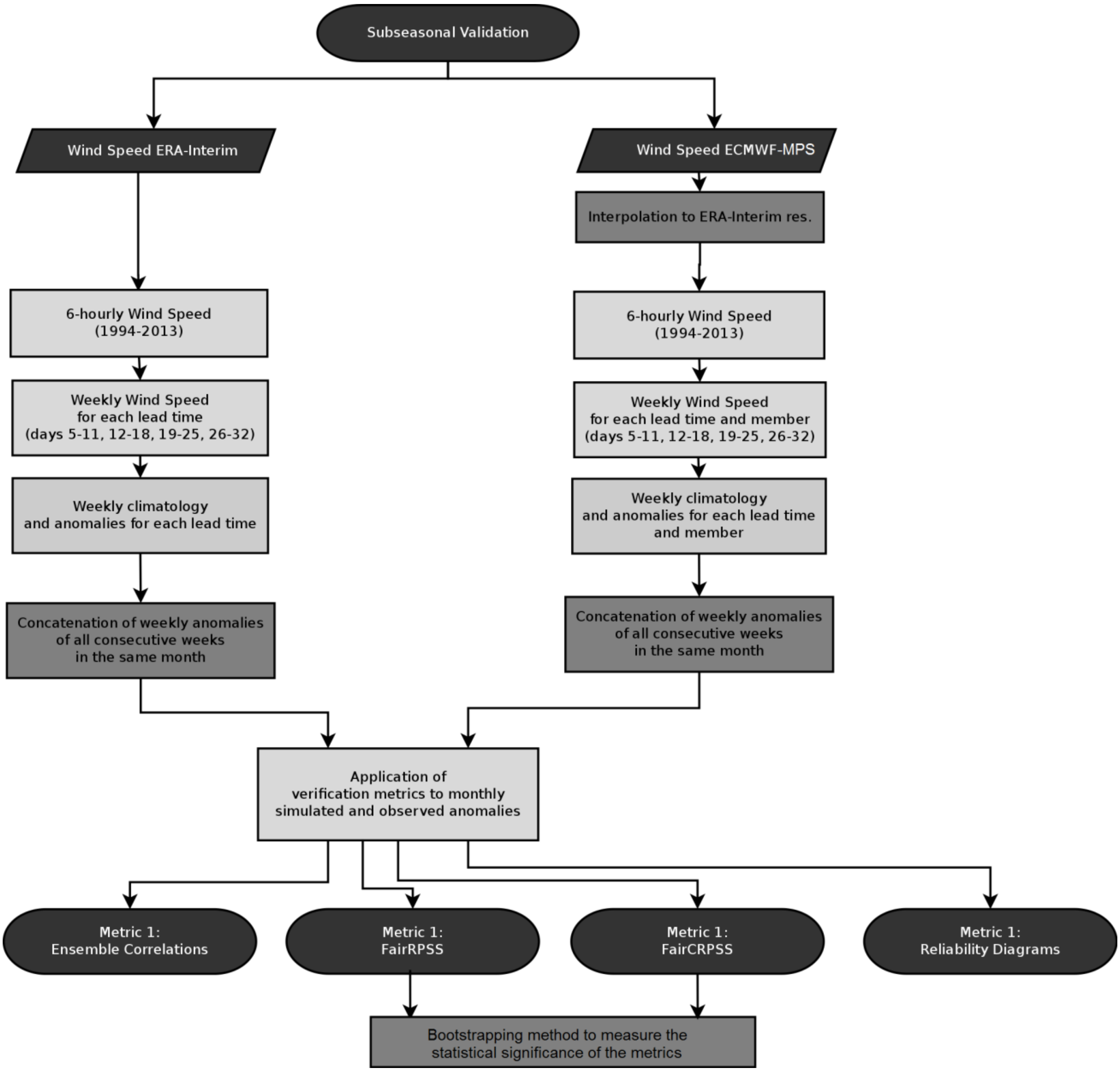

***Figure 1**. Flow chart of the quality assessment of the ECMWF-MPS.*

Flow chart with the sequence of the steps followed for the validation of the 10-metre wind speed forecasted with the ECMWF-MPS, including pre- and post-processes.

3. Results

3.1. Global quality assessment

Global maps of EnsCorr, FairRPSS and FairCRPSS are shown in Figure 2 for the five start dates of January (Thursday 2nd, 9th, 17th, 23th and 30th) and four possible lead times, corresponding to days 5-11, 12-18, 19-25, 26-32. It is evident that for all lead times, the EnsCorr has a higher skill compared to the FairRPSS or the FairCRPSS, since it represents the potential skill that the forecast system might.

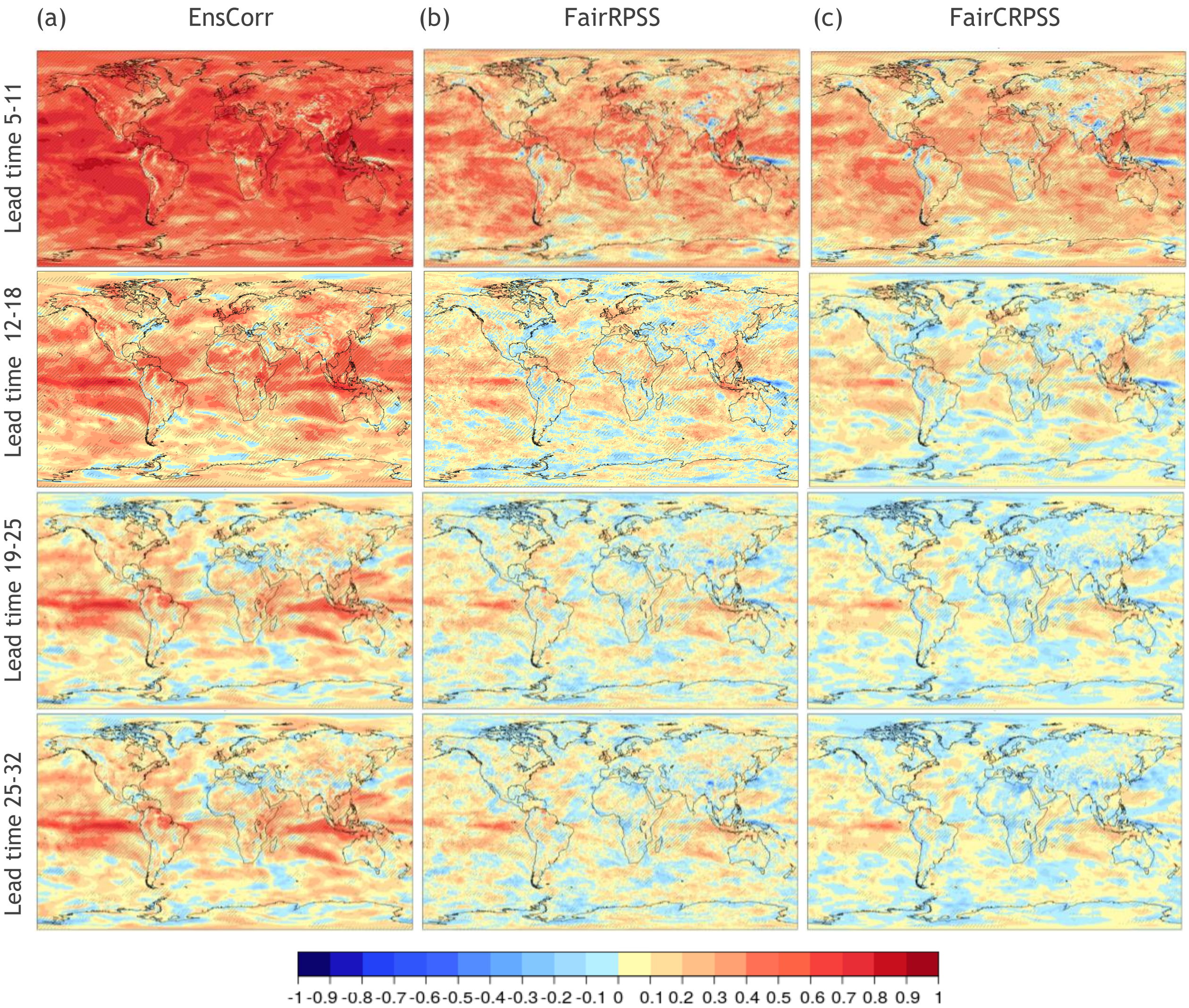

***Figure 2**. January skill scores.*

Skill scores for January start dates. (a) EnsCorr, (b) FairRPSS and (c) FairCRPSS. Each row shows a different lead time: days 5-11, 12-18, 19-25, or 26-21. Reference dataset: ERA-Interim (1994-2013).

By default, also the FairCRPSS is always slightly lower than the FairRPSS. Notwithstanding, the three scores show a spatial distribution globally similar (not considering both magnitude and significance). It is also clear that skill mainly decreases in the transition from the first lead time (days 5-11) to the second one (days 12-18). Skill decrease from the second to the third lead time (or from the third one to the fourth one) is considerably lower. Skill also exhibits seasonal variability, with a maximum in winter and a minimum in summer (see figures in the ESS web catalogue[1]). Such characteristics are typical not only of the start date of January but of all start dates.

If we focus only on the second lead time, we can make a few general remarks on January skill that are also mostly valid for the two subsequent lead times. Over oceans, EnsCorr are often significantly positive (at 95% confidence level), particularly along the tropical belt, and with many areas with correlations above 0.5 (Figure 3). Over continents, correlations are mostly positive and significant over North America, South America and Africa. Asia, Europe and Australia have mainly positive correlations, even though in many areas they are not significant.

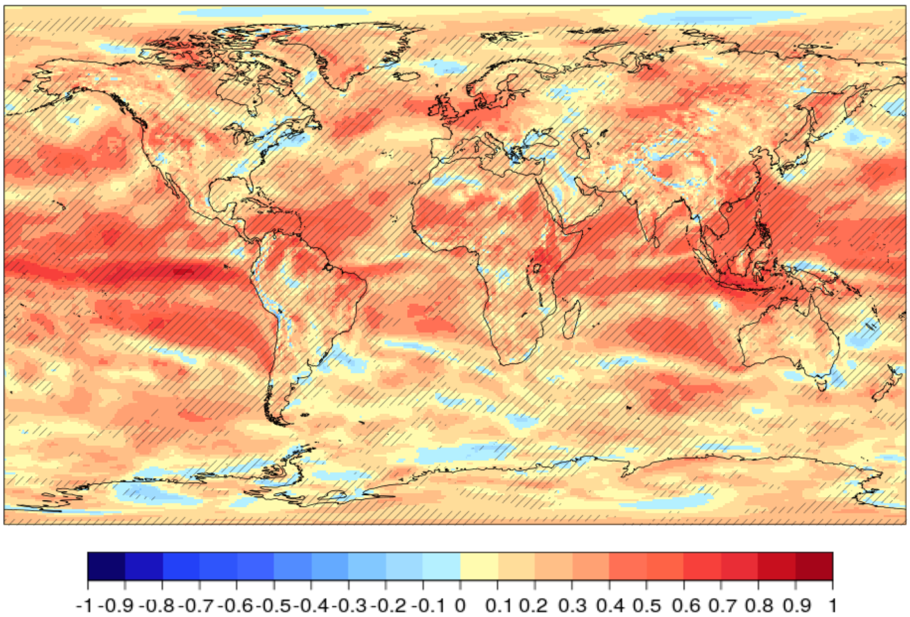

***Figure 3**. January EnsCorr for the second lead time (days 12-18).*

EnsCorr for January start dates and lead time 12-18 days. Dashed lines show areas where correlation is significant at 95% confidence level, obtained with a bootstrapping test. Reference dataset: ERA-Interim (1994-2013).

[1] www.bsc.es/ESS/catalogue

As expected, FairRPSS is much less significant and positive than EnsCorr (Figure 4). Over oceans, the highest FairRPSS, with significant values around 0.5, is still found along the tropics, particularly in the eastern Pacific Ocean, in the Indian Ocean and Indonesia. On the contrary, in the western Pacific Ocean close to Indonesia, a large area shows the highest negative skill measured for the whole globe (<0.5), even though these values are not significant. Atlantic Ocean is a region with moderate skill (usually above 0.2), especially near the northern part of South America and Caribes, where FairRPSS is often significant too, and in the North Sea. Over continents, significant values of FairRPSS show a less homogeneous spatial distribution. In North America fewer significant areas are observed, compared to South America. In Europe, England, Germany, Holland and Denmark illustrate positive skill, but it is rarely significant. In Africa, skill is significant in roughly half of the territory, without showing a clear spatial pattern. In Asia and Australia, FairRPSS doesn't show any significant extended region.

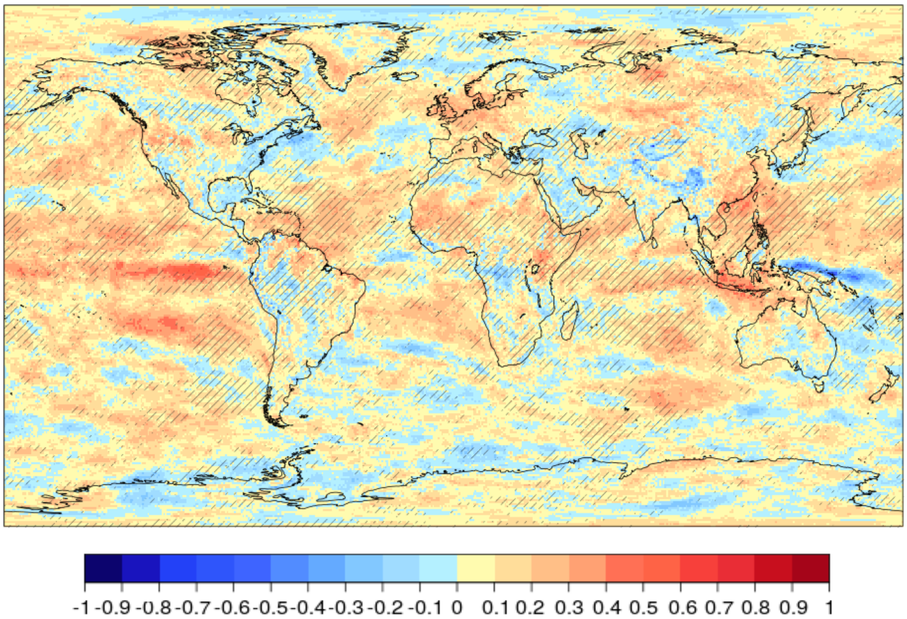

***Figure 4**. January FairRPSS for the second lead time (days 12-18).*

FairRPSS for January start dates and lead time 12-18 days. Dashed lines show areas where the skill is significant at 95% confidence level, obtained with a bootstrapping test. Reference dataset: ERA-Interim (1994-2013).

FairCRPSS has a total significant area similar to the FairRPSS, even if skill values are globally lower (Figure 5). Overall, spatial distribution of significative correlations is similar to the FairRPSS, both over oceans and continents.

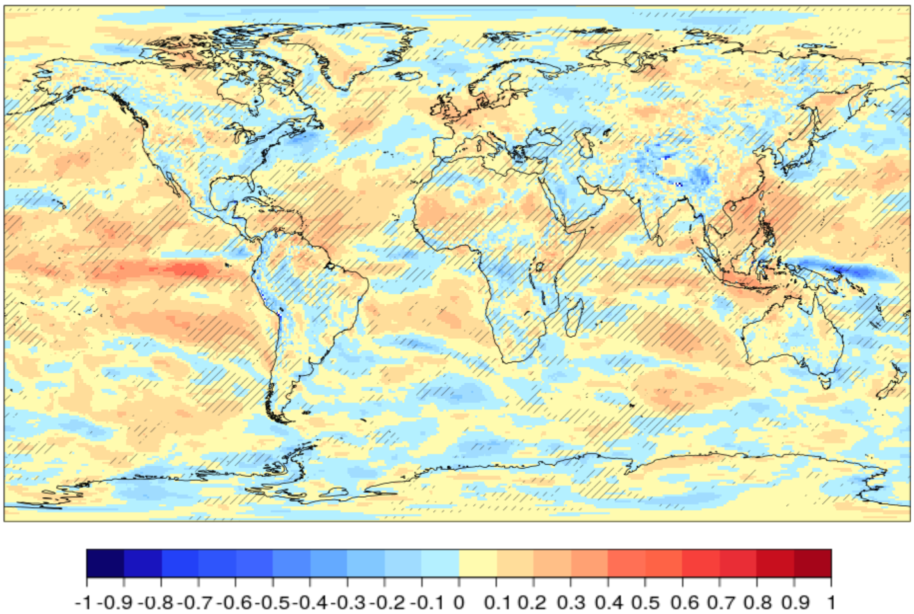

***Figure 5**. January FairCRPSS for the second lead time (days 12-18).*

FairCRPSS for January start dates and lead time 12-18 days. Dashed lines show areas where the skill is significant at 95% confidence level, obtained with a bootstrapping test. Reference dataset: ERA-Interim (1994-2013).

Globally, January monthly forecasts show that the areas with positive skill for the three verification measures are more extended than those with negative skill (even at high lead times), meaning that the monthly forecasts perform better than climatology in most part of the world. Verification measures for the other months of the year are shown in the ESS web catalogue[2].

3.2. Regional quality assessment

3.2.1. European region

Values of EnsCorr averaged over the European region (15°W-45°R, 35°N-75°N) are always greater than 0.5 from October to April (Figure 6, left) for the first lead time (days 5-11), with a maximum of 0.6-0.7 in November. They decrease to 0.3-0.5 from May to September. Correlations between the ensemble mean of the ECMWF S4 and ERA-Interim are considerably lower for other lead times different than the first one: highest values only reach 0.2-0.3 in December and February for the second lead time (days 12-18). At higher lead times, correlations are almost always positive but small (0-0.1). February is the only month when

[2] http://www.bsc.es/ESS/catalogue

EnsCorr is always over 0.1, even at high lead times.

Both FairRPSS and FairCRPSS, averaged over Europe, show similar values for all start dates and lead times (Figure 6, center and right), even if overall the FairCRPSS measures a skill slightly inferior than the FairRPSS, as it is expected by default (Wilks, 2011). Maximum FairRPSS values are measured from September to June for the first lead time (days 5-11), with a maximum in November (0.3-0.4). Lead time three and four (days 19-25 and 25-32) sometimes have negative FairRPSS (worse than climatology). Thus, ECMWF-S4 shows no little or no skill over climatology beyond the first lead time.

Reliability diagrams for all lead times and the central month of each season are shown in Figure 7. The five points shown in each diagram correspond to the forecast probability, which can have only five possible values: 0% if none ensemble members predict the tercile, 25% if one of the four members predict the tercile, 50% if two members predict it, 75% if three members predict it or 100% if all the four members predict it. It is evident from Figure 7 that for both upper and lower terciles, that the forecasts present a conditional bias, because they systematically underestimates events with small forecast probabilities (0% or 25%), and at the same time they overestimates all events with large forecast probability (50%, 75% or 100%), especially at lead times greater than 5-12 days. All months show a similar conditional bias, and the only intermonthly variability is observed for the last forecast class of 100%.

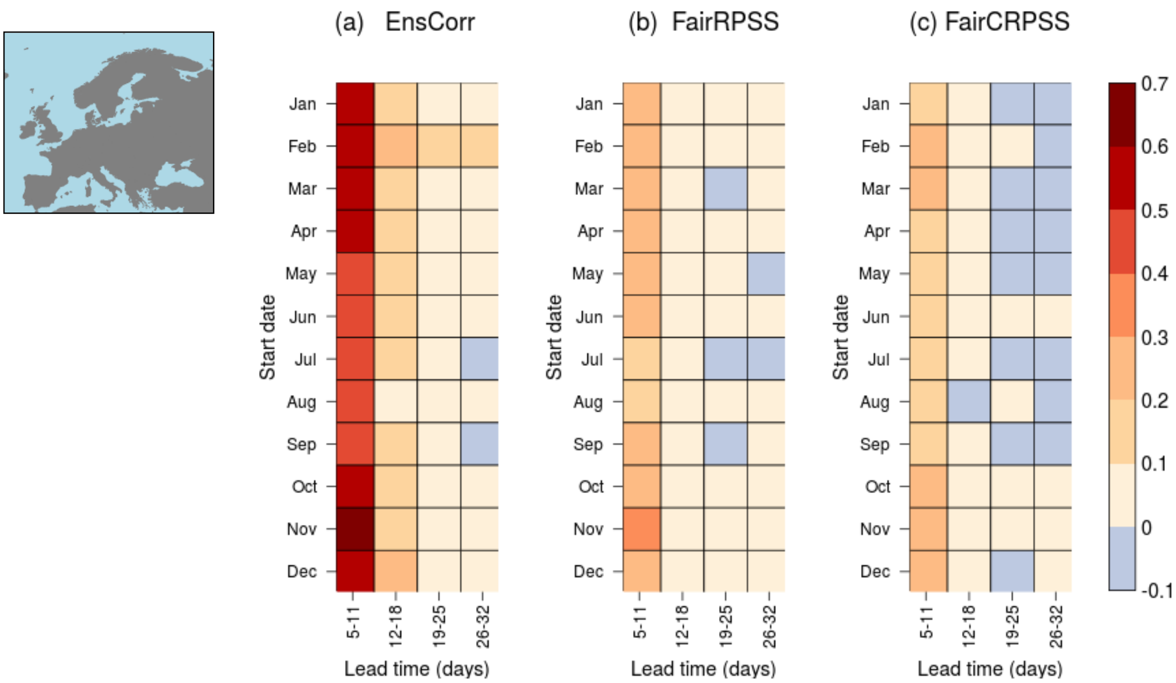

***Figure 6**. Skill scores for the European region.*

From left to right: EnsCorr (a), FairRPSS (b) and FairCRPSS (c) for each start date and lead time averaged over the European region (15°W-45°R, 35°N-75°N). Reference dataset: ERA-Interim (1994-2013).

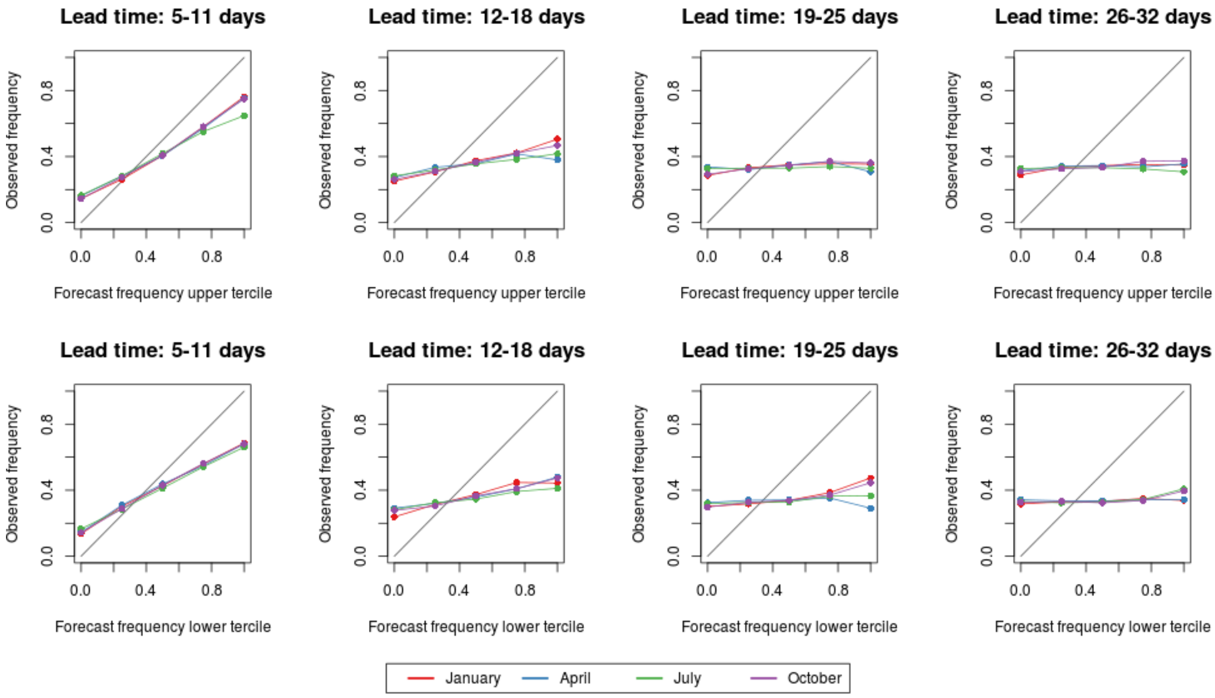

***Figure 7**. Reliability diagrams for the European region.*

Reliability diagrams of the upper (top) and lower (bottom) terciles for all the four lead times and four monthly start dates of January, April, July and October (the central month of each season) measured for the European region (15°W-45°R, 35°N-75°N). Reference dataset: ERA-Interim (1994-2013).

3.2.1. North Sea

Average ensemble correlations over the North Sea are shown at left in Figure 8. During the first lead time (days 5-11), EnsCorr is high (>0.5) from September to April, with a maximum from November to March (>0.6) while in the other months it is only slightly smaller (0.4-0.5). Ensemble correlations are lower during the other lead times: only December and January of the second lead time (days 12-18) have correlations above 0.3. February shows a constant skill between 0.2 and 0.3, even at high lead times. Negative correlations are observed in August for all lead times except the first and also during other months, but never for the first two lead times. Globally, compared to the European region (Figure 6), EnsCorr over the North Sea are slightly higher during winter months and slightly lower during summer months.

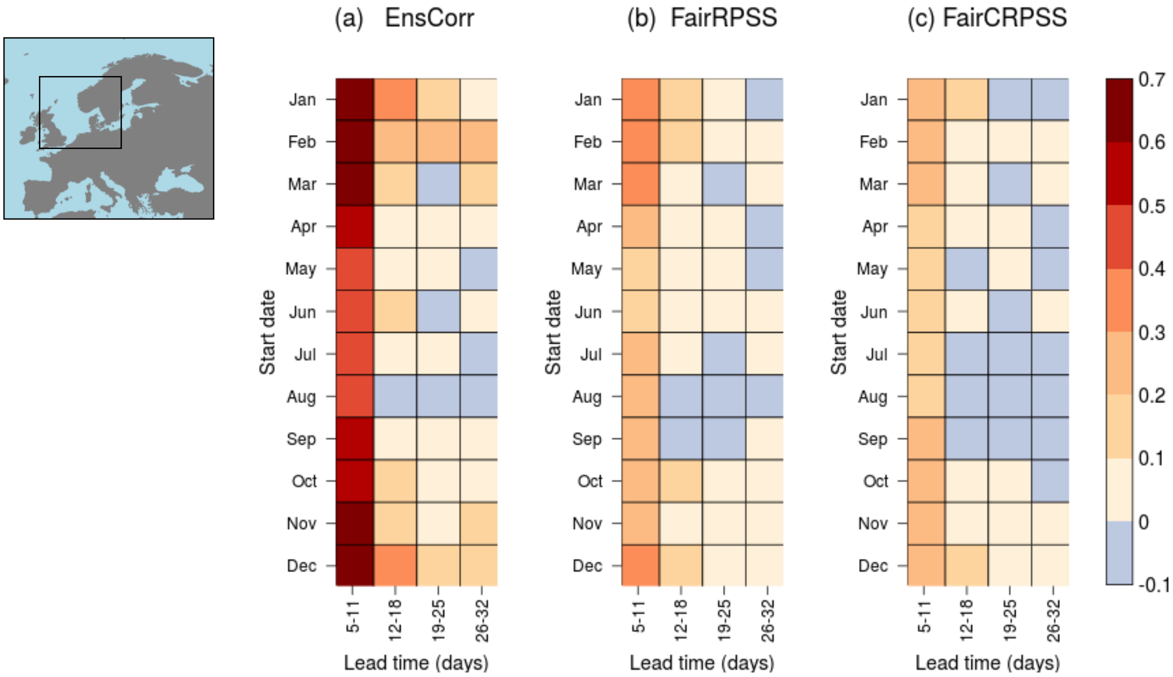

***Figure 8**. Skill scores for the North Sea.*

As Figure 6, but averaged over the North Sea region (4°W-15°R, 50°N-65°N).

A similar behaviour can also be observed for the FairRPSS and the FairCRPSS (Figure 8, center and right: skill varies seasonally, with a winter maximum and a summer minimum, and it is capped at 0.4 for FairRPSS and at 0.3 for FairCRPSS during the first lead time (5-11 days). Skill degrades for subsequent lead times, even if the forecasts keep being better than climatology (positive skill values), except during summer months and sometimes also outside summer for high lead times (days 19-25 or 26-32). Compared to the European region (Figure 6), the FairRPSS obtain higher values from December to March, but only during the first two lead times (5-11 and 12-18 days), while the FairCRPSS only shows a modest improvement over the same period and lead times. However, negative skill values are more frequent for both FairRPSS and FairCRPSS, especially during summer months.

Reliability diagram for the North Sea is shown in Figure 9. Conditional bias are similar to those observed for the European region, even if there is a higher intermonthly variability.

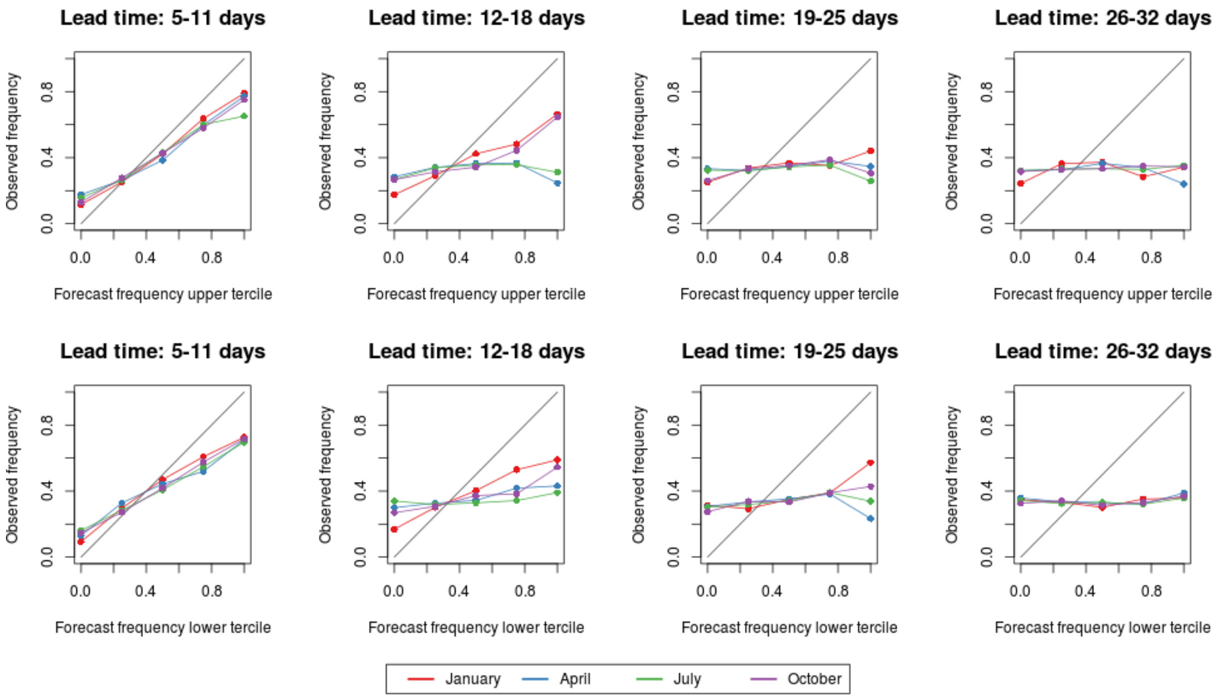

***Figure 9**. Reliability diagrams for the North Sea.*

As Figure 7, but for North Sea.

3.2.2. Iberian Peninsula

Average EnsCorr for the Iberian Peninsula are shown in the left part of Figure 10. Very high values (0.6-0.7) are observed from October to March for lead time 5-12 days, while other months of the same lead time range from 0.4 to 0.6. Other lead times have lower correlations: 0.1-0.4 for lead time 12-18 and up to 0.1 for lead times 19-25 and 26-31. February is the only month with average EnsCorr above 0.3 for lead time 12-18, and in general is the month with the highest potential skill, followed by April. Negative correlations sometimes are measured for lead times 19-25 and 26-32. Compared to the average European EnsCorr (Figure 6), correlations over the Iberian Peninsula are much higher from October to March (except in November), but only for the first lead time.

Both the FairRPSS and the FairCRPSS show moderate skill during the first lead time (0.3-0.4 during NDJ and MA months for FairRPSS and during DJ for FairCRPSS), and lower skill at higher lead times (up to 0.2), with negative skill measured during some months at high lead times, except July, which has negative FairRPSS even at lead time 12-18. Overall, the skill of the Iberian Peninsula, as measured by the FairRPSS and FairCRPSS, shows a similar interannual variation as the European skill (Figure 6) but with higher maximum values, even if limited to the first lead time. Reliability diagrams shown in Figure 11 reveal conditional bias similar to those of the North Sea (see Figure 9).

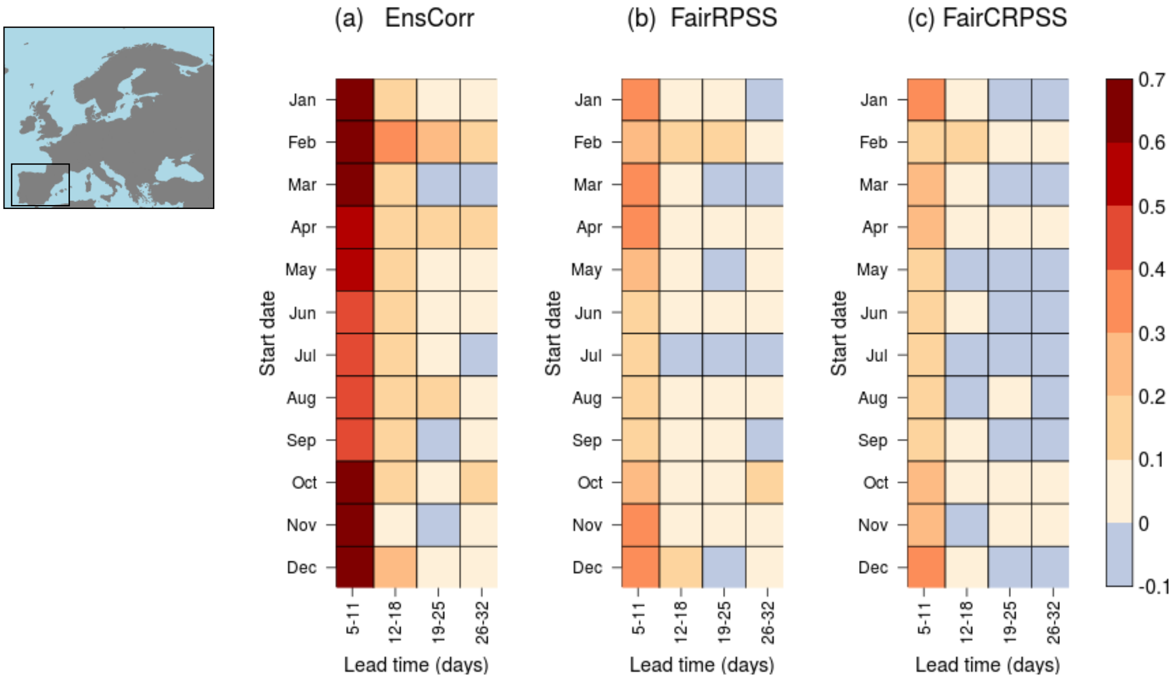

***Figure 10**. Skill scores for Iberian Peninsula.*

As Figure 6, but averaged over the Iberian Peninsula (10°W-4°E, 36°N-44°N).

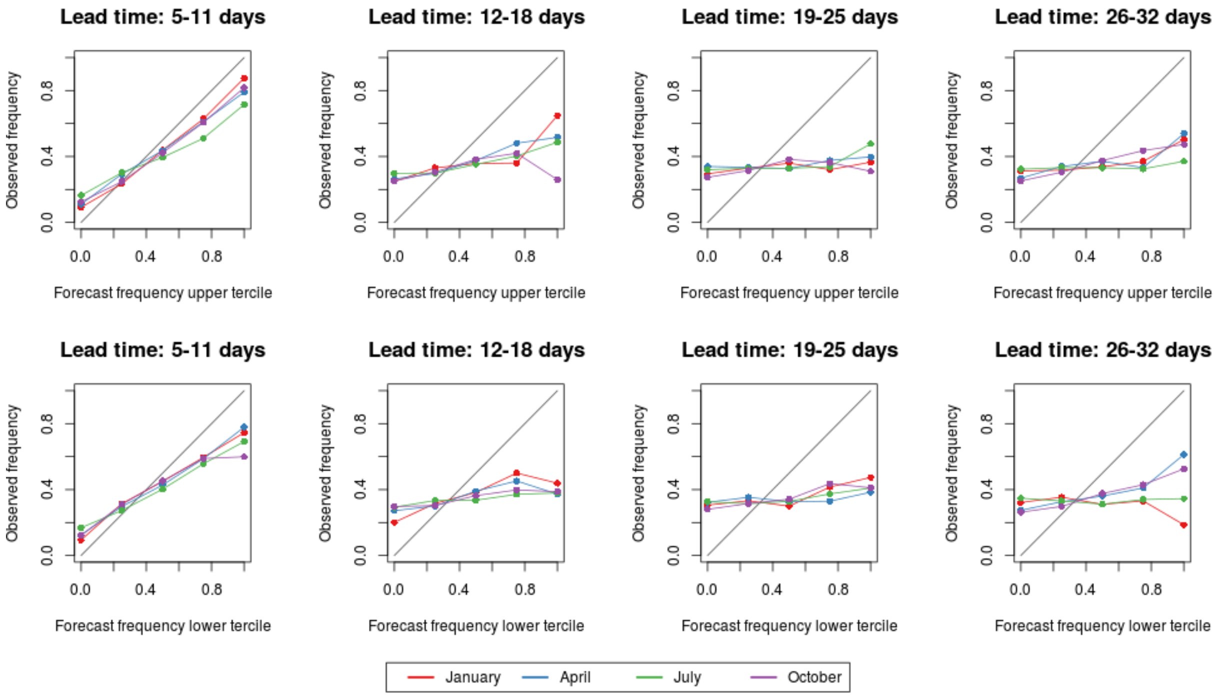

***Figure 11**. Reliability diagrams for the Iberian Peninsula.*

As Figure 7, but for Iberian Peninsula.

3.2.3. North America

Average EnsCorr for North America is shown in Figure 12, left. Ensemble correlations are high (>0.5) from October to June of the first lead time (5-11 days), with a maximum of 0.6-0.7 in December. Other months have EnsCorr between 0.4-0.5. Second lead time (12-18) always has positive EnsCorr and between 0.1-0.2, except in December and February (0.2-0.3) and in September (0-0.1).Third lead time (19-25 days) always shows a positive EnsCorr, which is higher in January and February (0.1-0.2). Fourth lead time (26-32 days) has negative EnsCorr only in June and September. Compared to the European EnsCorr (Figure 6), North America EnsCorr has higher skill during spring months of the first lead time (5-11 days), while correlations for the other lead times and months are remarkably similar to the European ones.

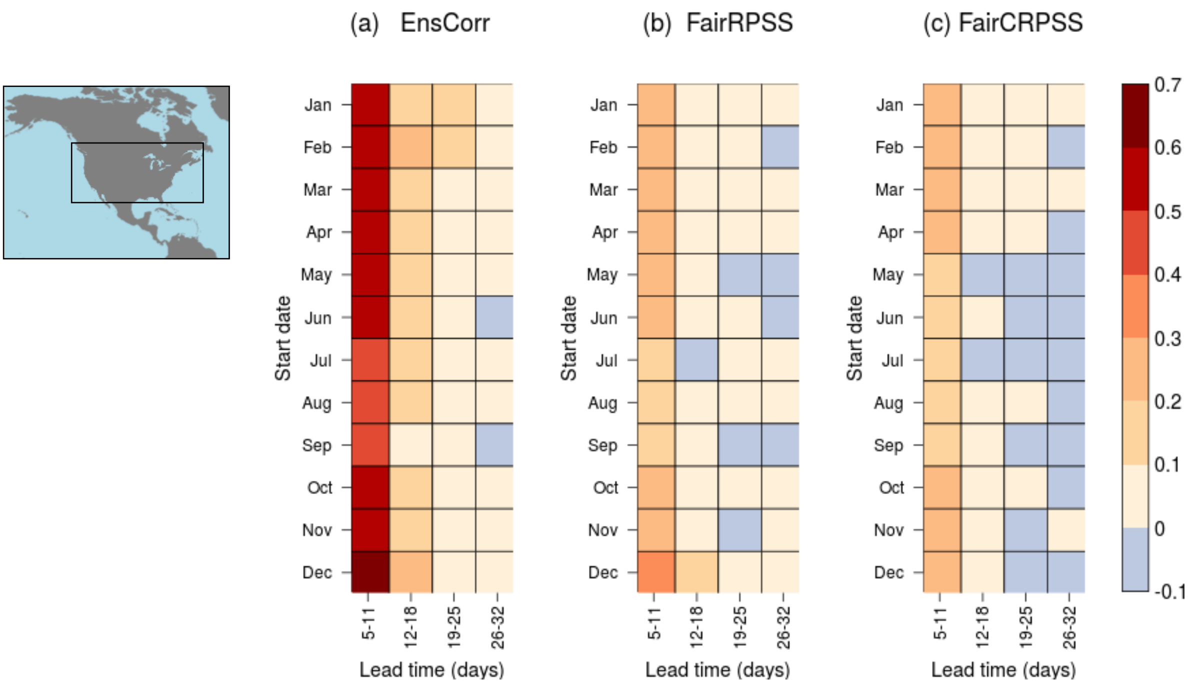

***Figure 12**. Skill scores for North America.*

As Figure 6, but averaged over North America (130˚W-60˚W, 30˚N-50˚N).

FairRPSS has a maximum value of 0.3-0.4 in December for the first lead time, while it is constantly above 0.2 from October to June and above 0.1 in the other months. Other lead times have FairRPSS values that don't exceed 0.1, but they are for the most part positive (forecasts better than climatology), On the contrary, FairCRPSS values are more negative than positive, except during the first lead time, when they are above 0.2 from October to April and above 0.1 in the other months, and during the second lead time, when they are above 0 during all months except May and July.

Figure 13 shows the reliability diagram for North America. Its conditional bias is of the same

type of that described for Europe (Figure 7). Measuring reliability over a larger area (compared to Iberian Peninsula or North Sea) reduces the differences between curves.

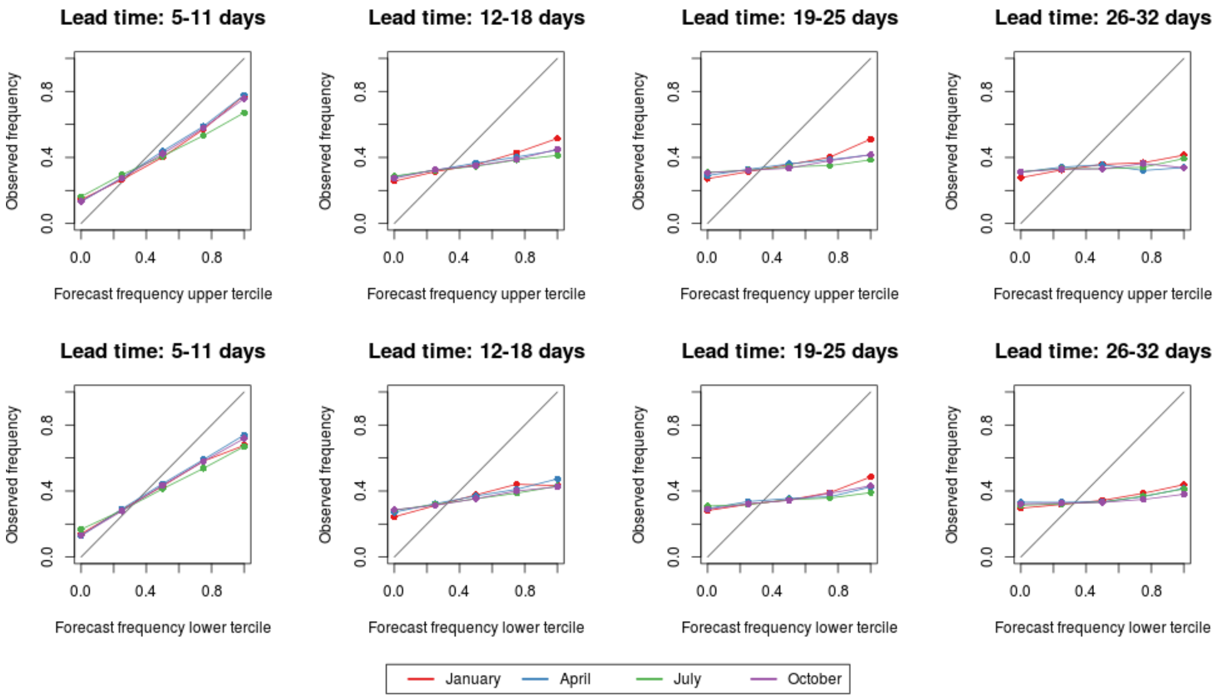

***Figure 13**. Reliability diagrams for the North America.*

As Figure 7, but for North America.

4. Discussion and conclusions

This report was developed in the framework of the RESILIENCE project (objective O1d3), and provides an exhaustive description of the predictability of the ECMWF monthly forecast system in simulating 10-m wind speed. For each monthly start date and weekly lead times, the wind speed forecasts are firstly assessed comparing them with observations from ERA-Interim reanalysis, employing one deterministic verification score and three probabilistic scores, respectively the correlation coefficient of the anomalies (EnsCorr), the fair ranked probability skill score (FairRPSS), the fair continuous ranked probability skill score (FairCRPSS) and the Reliability Diagram. A novel technique was introduced to be able to assess both monthly verification scores from weekly start dates and to increase the robustness of the validation. Four different key regions, crucial for generation of wind power, were selected and employed to measure the average of verification scores over the chosen area: Europe, North Sea, Iberian Peninsula or North America.

Results identified many regions with positive forecast skill, e.g. where monthly forecasts have higher predictability than climatology ('windows of opportunity') for several different lead times, usually during the winter half of the year. A number of reasons may explain the higher skill during winter months: larger SST gradients, stronger coupling between the stratosphere and the troposphere, and influence from the MJO (Lynch et al. 2014). Skill decreases particularly in the transition from the first lead time (days 5-11) to the second one (days 12-18), and it is usually highest and more significant around the equator region and all tropics in general, especially in the eastern part of the Pacific and in the Indian Ocean.

Lead time of 5-11 days, in particular, always shows positive skill in all regions considered, up to a maximum EnsCorr of 0.7. Focusing only on subsequent lead times, and on regions with FairRPSS >0.1 (to select only the strongest windows), it is interesting to notice that North America presents a good windows of opportunity in December during the second lead time, while Iberian Peninsula presents four, one in December for lead time 12-18, two in February for lead times 12-18 and 19-25, and one in October for lead time 26-32. The last window is the only one detected with FairRPSS above 0.1 and lead time 26-32, for all the regions considered in this study. North Sea also presents four windows of opportunity with FairRPSS > 0.1, but they are all circumscribed to the second lead time (12-18 days). European region is the only one that doesn't present any month and lead time (beyond the first) with FairRPSS > 0.1. Detailed figures of the skill for each start date and lead time are available in the ESS web catalogue[3].

Globally, wind speed monthly forecasts perform better than climatology in most part of the world. Sometimes, a negative FairRPSS and/or FairCRPSS can be observed, especially during summer months, and more frequently at higher lead times. It is worth mentioning that even when FairRPSS and/or FairCRPSS is negative (i.e: climatology is better than forecasts), the average of the skill over the study region never drops below the value of -0.1, meaning that

[3] http://www.bsc.es/ESS/catalogue

the forecasts are never much worse than climatology. Finally, reliability diagrams detected a conditional bias that underestimates low-probability forecast probabilities and overestimates high-probability ones, for all start dates, lead times, terciles and regions considered.

5. Acknowledgements

The authors acknowledge funding support from the RESILIENCE (CGL2013-41055-R) project, funded by the Spanish Ministerio de Economía y Competitividad (MINECO).

6. References

Dee, D.P. et al., 2011. The ERA-Interim reanalysis: Configuration and performance of the data assimilation system. *Quarterly Journal of the Royal Meteorological Society*, 137(656), pp.553–597.

Ferro, C.A.T., 2014. Fair scores for ensemble forecasts. *Q.J.R.Meteorol.Soc.*, 140(July), pp.1917–1923.

Foley, A.M. et al., 2012. Current methods and advances in forecasting of wind power generation. *Renewable Energy*, 37(1), pp.1–8. Available at: http://dx.doi.org/10.1016/j.renene.2011.05.033.

Hartmann, C.. et al., 2002. Confidence builders: evaluating seasonal climate forecasts from user perspectives. *American Meteorological Society*, (May), pp.683–698.

Hudson, D. et al., 2011. Bridging the gap between weather and seasonal forecasting : intraseasonal forecasting for Australia. *Quarterly Journal of the Royal Meteorological Society*, 1989(April), pp.673–689.

Joliffe, I.T. & Stephenson, D.B., 2011. *Forecast Verification - A practicioner's guide in Atmospheric Science* 2nd ed. WILEY, ed.,

Lynch, J., Brayshaw, J. & Charlton-Perez, A., 2014. Verification of European Subseasonal Wind Speed Forecasts. *American Meteorologica society*, (2006), pp.2978–2990.

Mason, S.J., 2008. Understanding forecast verification statistics. *Meteorol. Appl.*, 40, pp.31–40.

Molteni, F. et al., 2011. *The new ECMWF seasonal forecast system (System 4)*, European Centre for Medium-Range Weather Forecasts.

Robertson, A.W. et al., 2015. Improving and promoting subseasonal to seasonal predictions. *Bull. Amer. Meteor. Soc.*, (March), pp.49–53.

Rodwell, M.. & Doblas-Reyes, F.J., 2006. Medium-Range , Monthly , and Seasonal Prediction for Europe and the Use of. *American Meteorological Society*, 19, pp.6025–6047.

Vitart, F., Medium, F. & Weath, R., 2015. Evolution of ECMWF sub-seasonal forecast skill scores over the past 10 years. *Quarterly Journal of the Royal Meteorological Society*, (2013).

Weigel, A.P., Baggenstos, D. & Liniger, M.., 2008. Probabilistic Verification of Monthly Temperature Forecasts. *American Meteorological Society - Monthly Weather Review*, 136, pp.5162–5182.

Wilks, D.S., 2011. *Statistical Methods in the Atmospheric Sciences* 2nd ed., International Geophysics Series.